节律在听觉言语理解中的作用

陈梁杰 刘雷 葛钟书 杨晓东 李量(通讯作者)

心理与认知科学学院，北京大学，北京，中国

通讯作者邮箱：liliang@pku.edu.cn

节律在听觉言语理解中的作用

摘 要 言语理解是听者接受外部语音输入并且获得意义的心理过程。日常交流中，听觉言语理解受多尺度节律信息的影响，常见有韵律结构节律、语境节律、和说话者身体语言节律三方面外部节律。它们改变听者在言语理解中的音素判别、词汇感知以及言语可懂度等过程。内部节律表现为大脑内神经振荡，其能够表征外部言语输入在不同时间尺度下的层级特征。外部节律性刺激与内部神经活动的神经夹带能够优化大脑对言语刺激的处理，并受到听者自上而下的认知过程的调节进一步增强目标言语的内在表征。我们认为它可能是实现内外节律相互联系并共同影响言语理解的关键机制。对内外节律及其联系机制的揭示能够为理解言语这种在多层级时间尺度上具有结构规律的复杂序列提供了一个研究窗口。

关键词 节律，言语理解，神经振荡，神经夹带，自上而下调节

节律在听觉言语理解中的作用

1 引言

从生命更迭到昼夜交替，自然界中充溢着繁杂的节律变化。击鼓、起舞或歌唱，这些活动中的拍手，踏步或发声通常按照一定的周期循环往复。在人类演化进程中节律无处不在，它们承载着信息传递的重要作用(Kotz et al., 2018)，是社会交流和互动的重要媒介。长期以来，节律研究主要关注于感知觉加工，忽视了其在更为复杂的言语理解中的作用。直到近期，研究者开始逐渐重视节律对言语理解的影响，并通过记录听者内部的神经活动揭示节律影响的作用机制。

节律的内在时间组织调节着个体间沟通和互动方式，口头言语作为人类社会中重要的信息交流渠道具有丰富节律特性。在判定对象是否具有节律特性时存在两种方式，一种强调时间上规律性，另一种强调时间上的结构关系。前者定义的节律可以被称为协调节律或周期性节律，它意味着固定间隔和模式的不断重复 (White, 2014; White et al., 2012)。例如，钟表转动过程中发出的“滴答、滴答”声响和正常心脏的起伏跳动，这些都具有时间上的规律性或近似规律性。而言语的节律性更偏向后者的定义，即一个给定的属性或属性间的组合在一段时间跨度中的稳定关系(Fujii & Wan, 2014; Ramus et al., 1999)，例如，树干中的硬结让锯木出现卡顿，但我们仍旧会认为伐木工来回的动作具有节律性。

在语言学领域中，早期对西班牙语中“机关枪”式，英语的“莫尔斯电码”式与日语发音的感知使研究者关注于语系间不同言语层的等时性现象，据此将节律感知划分为重音计时（stress-timed）、音节计时（syllable-timed）和亚音节（mora）三类形式（Ladefoged, 1975; Pike, 1945; Ramus et al., 1999）。但这种分类方式过于强调单元间的等时性，在多语系的语音信号分析中无法为 “等时理论”提供经验支持（Dauer, 1983; Ramus, 2002; Roach, 1982）。后来基于元音持续时间变化的分类方法更具有实证性，它依照口语中元音所占时间比例的不同试图建立起一种更为广泛的节律量化方式（Ramus et al., 1999）。如重音计时相比音节计时语系，元音持续时间更加多变（Ling et al., 2000）。这些分类方式说明言语相比于单一振荡器形成的特定间隔的重复活动并不具有客观的等时性的周期特点，但仍可以被直观地感知为节律活动（Jadoul et al., 2016; Kayser, 2019; Turk & Shattuck-Hufnagel, 2013）。与音乐中的节奏感相近的是，言语中的单个属性，如音调变化或音节持续时间长短，也能带来的主观上的节律感受（Dellwo, 2006）。但只专注于个别属性的度量并不能囊括言语节律的所有特征，言语的节律感知还取决于一系列其他因素，包括整体响

度变化和语速高低等(Nooteboom, 1997)。这些因素共同作用于听者的知觉加工过程，使听者感受到言语中的节律性。

言语理解是听者根据外部言语输入（如目标语音）和背景信息（如语境或非言语信息）获得意义的心理过程(杨玉芳，2021)，包括音节，词汇和句子不同层级的加工(Farbood et al., 2013; Sheng et al., 2019)。说话者口语在韵律结构上具有一定的节律特性，如重音位置和语速快慢等，这些节律的改变会影响听者对目标语音的理解水平。当说话者的说话语速或音节产生速率超出正常范围（3 - 8 Hz）时，言语可懂度会出现显著的降低(Ahissar et al., 2001)。相比之下，背景语境节律则会改变听者的音节层面感知，如事先呈现一串有规律的纯音序列，其呈现速率的不同会改变个体感知后续出现的辅音，即节律较快的纯音序列会使被试将辅音更多知觉为/w/而不是/b/（Wade & Holt, 2005）。背景信息不单表现在声学层面的变化上，由于言语语音的时域包络、发声者的声道活动与肢体动作三者之间是高度关联的，这决定了影响言语理解还涉及非言语节律，如说话者的身体语言（body language）。身体语言包括面部运动、身体姿势、手势等活动(Müller et al., 2013)，说话者的面部运动往往与言语的时域包络起伏具有相似的节律特性，这有助于听者更好理解言语信息（Ghazanfar & Takahashi, 2014）。据此本文提及的外部节律定义为听觉言语理解过程中能够对言语理解产生影响的客观世界中具有节律特征的物理输入。我们将围绕言语理解涉及的言语输入和背景信息所包含的三类常见外部节律，分别为韵律结构节律、语境节律和说话者身体语言节律，阐述其在言语理解中音素、词汇和句子三个层级的影响，借此说明外部节律对言语理解的作用。

听者大脑是如何利用外部节律促进或改变言语理解的呢？这一过程被认为和内部节律即神经振荡（neural oscillation），一系列颅内神经元集群节律性地、同步性地电活动，存在明显联系（Haegens & Golumbic, 2018; Kösem et al., 2016; Zion-Golumbic & Schroeder, 2012）。神经振荡作为皮层神经元集群活动被认为介导了不同的认知加工过程，包括言语加工，抑制干扰等（Jensen et al., 2012; Schroeder & Lakatos, 2009）。近期研究认为内部节律活动可能受外部节律所影响，表现出内外节律随时间推移相趋近的现象（Lakatos et al., 2019; Obleser & Kayser, 2019），这种现象被称为神经夹带（neural entrainment）。当内部节律与外部目标言语发生夹带时，听者表现出更好的言语理解成绩（Riecke et al., 2018; Vanthornhout et al., 2018）。同时，言语理解的多种高级认知过程同样能够调节神经夹带的表现，如选择性注意（Arnal & Giraud, 2012; Helfrich et al., 2019; Lakatos et al., 2013）、先验语法知识（Ding et al., 2016; Ding et al., 2017）和语境预期（Broderick et al., 2019）等。

基于此，我们认为神经夹带可能是言语理解过程中实现内外节律相互联系与共同作用的关键机制。

综上，本文首先论述三种常见的外部节律如何影响听觉言语理解，说明节律对言语理解影响的普遍性。接着，我们总结了作为内部节律的神经振荡在言语理解中的功能。最后结合神经夹带在言语加工中的作用与其受自上而下认知过程的影响，讨论神经夹带在言语理解中作为联系内外节律机制的可能性。未来研究需要从不同层级、不同尺度和不同背景中去探究节律在听觉言语理解中的意义。

2 外部节律与言语理解

言语产生是随着时间推移而展开的，这使得时间规律对听者理解信息至关重要。为了理解言语内容，听者需要根据外部节律特征从持续的语音流中感知音素、音节、单词和短语的时间组织（Ding & He, 2016; Kotz & Schwartze, 2010; Peelle & Davis, 2012）。此部分我们依据言语输入和背景信息将影响言语理解的外部节律划分为三种常见类型，主要涉及韵律结构节律、语境节律和说话者身体语言节律。

2.1 韵律结构节律改变句子可懂度

言语的韵律结构节律在阅读和口语交流中表现各不相同。视觉阅读中词与词组合而形成的音节数目的搭配会动态影响局部短语分析和整体句子整合（Luo, Y & Zhou, 2010; Luo, Y et al., 2015），视觉输入为主的阅读并不能直接提供韵律结构信息而需要读者借助内部表征如默读等方式实现。此部分主要关注于听觉场景中外部节律的韵律结构节律，即口语中的音节长短、音节间间隔和重音分布等特征上（Dellwo, 2006; Ramus et al., 1999; 方岚 et al., 2021）。

音节间间隔能够直接影响言语可懂度。研究者通过对句子进行时间上的压缩降低了音节间停顿时间，导致整体语速加快，结果发现听者对句子的可懂度出现了剧烈的下降（Bosker & Ghitza, 2018; Ghitza & Greenberg, 2009）。听者表现出难以加工韵律结构节律被破坏的言语刺激，但这可能是源自于句子加工依赖于特定节律的感觉输入，又或者时间压缩后音节内的声学结构被破坏使听者难以识别。为了回答这一问题，研究者将压缩后的言语波形进行等距的切分，切分后每段内的音节仍旧处于压缩状态，而后在每个片段后加入一段无声的间隔以产生人为的节律特性。听者加工这类句子的可懂度得到了恢复。需要注意的是，只有在插入的间隔是以固定规律形式时言语可懂度才会恢复，无规律的间隔则没有任何效应（Ghitza & Greenberg, 2009）。所以听者对于言语的理解依赖于言语自身节律特性，在压缩后的句子中加入无声间隔的过程可以理解为是对句子内音节的“重新打包”

（repackage），即将时域上的波形分割成不同部分。这些包以规定的速率传导至双耳，协助听者预测包内音节的最大信息传输速率从而在一定程度上恢复了言语可理解性。

韵律结构节律对可懂度的影响反映了听觉系统在处理不同传输速率的信息流时的自适应性。自然言语中，音节间停顿的长短主要涉及到两个方面因素（Ghitza & Greenberg, 2009），一个是人类发音器官的生物力学特性和大脑的神经动力学特性。发音器官/大脑内在振荡导致口唇运动和言语的时域包络大约为 5 Hz 的节律，以此调制沉默时间的长短；另一个因素是言语的层次韵律结构。例如，当一个音节在一个单词内时，它后面的停顿期通常很短，但当它与一个更高层的语言结构(例如韵律词、韵律短语和语调短语)的边界重合时，沉默会逐渐拉长。言语中的沉默给大脑提供了额外的时间来处理间隙之前的音节，当沉默时间被缩小或扩大并违反自然语言的时间规律时，会增加听者加工所需的负载，进而破坏句子的可懂度(Ding & He, 2016)。

除了停顿的长短外，停顿的位置也被认为能够改变听者对于言语的节律感知，主要涉及口语句子中韵律边界（prosodic boundary）。这些边界与感知停顿、边界前音节延长和短语末尾的音高都具有联系(Li & Yang, 2009, 2010)，所以韵律边界的感知能够促进听者将言语切分成不同层次的组块，并与口语的感知流畅性和可理解性密切相关(方岚等人，2021；于泽等人，2010)，汉语作为声调言语在口语的结构分析、语义加工和情绪感知等方面均受到韵律边界的影响(Li & Yang, 2010; Li et al., 2019)。近期研究以汉语中同时可理解为偏正结构（modifier noun construction）或述宾结构（narrative object structure）的歧义短语为材料发现，当听者注意到韵律信息时，韵律边界会改变听者对歧义短语的结构分析过程(Li et al., 2019)。韵律边界能够在歧义语境下帮助听者进行句式结构的分析，通过消除结构上的歧义从而促进言语可懂度。此外，汉语韵律中的重音位置还能够改变听者对言语中不同位置词汇的选择性注意，使得重音后的词汇被更强的加工(Li & Ren, 2012)。

2. 2 语境节律改变词汇与音素感知

目标言语前后的声学场景我们一般会称为语境（context），可以在时间上与目标语音相邻或不相邻(Stilp, 2020)。语境对言语理解的影响主要源于语速快慢，这是因为在词汇感知或边界切分中听者需要依赖语境中提供的相对速率线索(Brown et al., 2011; Dilley et al., 2010; Dilley & McAuley, 2008)。说话者语速变化一般通过句子中元音和辅音间间隔调节来实现(Dellwo, 2006)，而这两类音素间间隔的分布能够反应语句中节律性(Ling et al., 2000; Ramus et al., 1999)。当说话者的语速较慢时，听者容易将句子中所包含的一个虚词（如：or 或 are）忽略，即在知觉层面表现出词汇消失的现象(Dilley & Pitt, 2010)。有趣的是，如

果把语速加快听者反而会知觉到句子中本来没有的虚词。这种现象会随着语境时间拉长而加剧，Baese-Berk 等人(2014)同时操纵了全局语境（global-context，整段材料）的语速和远端语境（distal-context，目标词所在句子）的语速，发现随着时间推移全局语速对目标单词数量感知的影响增大，即随着语速的减慢单词数量感知呈现下降趋势。这些发现表明随着语境节律的加速或减慢，听者对言语中词汇数量的知觉会向补偿方向移动以确保感知保持稳定，即听者会自发调整对于后续出现词汇的持续时间或者边界位置的主观感知来契合整体语境节律。值得注意的是，这一现象可能特异于语境内容，人为降低语境内语音的可懂度或者采用其他音调序列时，听者对单词数量的识别将不受影响(Pitt et al., 2016)。

语速除了改变听者对语境下词汇数量的判断，还会影响单词内元音和辅音的辨识, 这是因为言语感知很大程度上依赖于从特定频率信息中恢复音位线索（Di Liberto et al., 2019; O'Brien et al., 2020）。例如，一个快速的语音环境会使听者更偏好于将一个模糊元音判断为长元音(如：/a, a:/)， 因为相邻语境内的音节持续时间会改变后续音节持续时间的主观评估，快节律的语境会使听者对客观时间的判断变短使后续元音听起来相对较长（Bosker, Sjerps, et al., 2020; Kosem et al., 2018; Reinisch, 2016）。语境速率对音素间语音边界的影响被称为语音边界移位(phonetic boundary shift, PBS)（Maslowski et al., 2019; Reinisch, 2016）。这种现象同样发生在辅音感知中，即较快节律的语境对/ba/-/wa/的模糊音节判断中，听者会更大可能性认为是/wa/（Wade & Holt, 2005）。言语中音素知觉受语境带来的外部节律影响，由于在非言语环境下这种现象同样会被诱发，如纯音序列（Bosker, 2017），所以这种速率依赖的知觉被认为涉及一般的听觉过程。

知觉从来不是对感觉信息的客观登记。就像任何形态的感知一样，言语感知是相对于语境的，它会依据先前的经验和背景发生改变（Stilp, 2020）。上述研究表明在外部节律的诱导下，听者对给定语境下感知到的单词数量和音节判别会发生改变。这些结果有助于解释在言语信号失真情况下，语音识别能力下降的原因。

2.3 身体语言节律对言语理解影响

身体语言是一种非语言交流模式，说话者通过面部活动和手部摆动等方式辅助自身的信息表达(Holler & Levinson, 2019)。在面对面的交流中听者同时感知到的说话者的身体活动和言语节律往往在特定频率上匹配，这有利于它们之间的耦合，就像“手舞足蹈”的演讲者会更容易让听众关注于演讲内容（Morillon & Baillet, 2017; Morillon et al., 2014; Rimmele et al., 2018）。

说话者发音器官内的一系列协作运动体现在声道的开放和缩小循环中（Abbs et al., 1984; Browman & Goldstein, 1992; Cho et al., 2019; Proctor et al., 2019）。例如，在发音/b/时需要封闭声道的前部，因此口唇和下巴的运动之间有一个协同过程，以实现完全闭合。目前许多研究都集中在言语中的声音和运动的相互作用上，如听者在观测说话者的口唇运动时，通过人为改变运动速率会影响听者对实际言语的语速判定（Bosker, Peeters, et al., 2020）。而在多说话人场景中，研究者还发现说话者的口唇运动信息能够提升听者对目标言语的识别成绩（Wu et al., 2013）。听者对言语的加工除了利用说话者的口唇运动，还会根据其自发的手部运动来理解言语（Iani & Bucciarelli, 2017; 殷融, 2020）。说话者为了表明言语中的重点往往会利用手势的摆动突显重音位置，研究者发现说话人的手部两相运动(上下摆动手臂)会显著改变听者对词汇中的重音感知位置（Bosker & Peeters, 2021）。这些结果都在说明言语知觉不仅受听觉层面外部节律的影响还受说话者的非听觉的运动节律的影响。

听者能够利用非声学的身体语言节律信息促进言语理解，可能说明听者和说话者之间存在某种重合的先验知识。言语知觉的运动理论（motor theory of speech perception）认为说话者和听者会共享一套相似的神经运动指令（neuromotor command），当听者加工说话者的运动信息并将其映射到自身的指令时，这将有助于听者理解说话者的言语内容（Poeppel & Assaneo, 2020）。

综上，外部节律对听觉言语理解的影响存在于广泛听觉与非听觉刺激中，语境语速能够改变听者对后续音素的判别和词汇数量的估计，言语内在节律能够改变句子可懂度的高低，身体语言节律可以改变重音位置感知。但是我们的大脑是如何利用这些节律信息指导言语感知，接下来我们将从神经元集群的节律性振荡方面进行讨论。

3 外部节律影响言语理解的神经机制

早期关于听觉言语理解的大脑内部过程研究主要采用事件相关电位（event-related potentials, ERPs）和功能磁共振成像（functional magnetic resonance imaging, fMRI）技术展开。音节探测、言语理解涉及 N1-P2、N400、P600 等事件相关电位成份的参与（Bridwell et al., 2018; Broderick et al., 2018; Morris & Klerke, 2016）。近年来随着研究方法的改进，通过颅内电极记录和时频分析等手段，大脑内自发的神经振荡开始成为关注的对象，从神经振荡层面揭示听觉言语理解的研究越来越多。本部分将关注言语加工中大脑的内部节律变化，以及神经夹带现象在其中的作用。

3.1 听者的内部节律——神经振荡

言语本身的节律特性作为其内在属性是如何在大脑中得以表征，以及外部节律如何影响言语感知？为了回答这些问题，研究者开始关注大脑内部节律活动的作用 (Ding et al., 2016; Zion-Golumbic & Schroeder, 2012; Haegens & Golumbic, 2018)。在早期，通过头皮记录到的电活动变化一直被认为是大脑活动的背景噪音，后来研究者开始意识到神经元集群的振荡活动体现了神经元兴奋性的周期变化(Bishop, 1933; Reichle, 2010), 如振荡的瞬时相位反映了神经集群在给定时刻的兴奋性水平（excitability level）。当振荡的兴奋性阶段被调整，使神经元集群高兴奋性与任务相关的感官输入相一致，对齐的输入将得到最优处理(Schroeder & Lakatos, 2009)，所以大脑的内部节律可能是完成外部节律性刺激加工的理想工具。

神经振荡依照频率高低的常被划分为 delta 频带（1-4 Hz）、theta 频带（4-10 Hz）、alpha 频带（8-15 Hz）、beta 频带（12-30 Hz）和 gamma 频带（30-200 Hz）。在听觉言语加工中，theta 频带振荡被认为能够将输入的连续语音信号分解为离散的单词单元，而 delta 频带振荡则把分割的单词结合为更高层的基于语法或语义组合的言语结构（Kösem et al., 2016; Ding et al., 2016）。近期在汉语的韵律语境加工中也发现，韵律节律可能通过增强与语音加工相关的频带活动促进语音理解。相比于不规则韵律节律的语境，规则韵律节律能够诱发听者在加工目标名词前的 beta 频带和目标名词后的 alpha 频带增强（Li et al., 2019）。而更高频的 gamma 频带的包络变化则被发现能够表征语音在功率谱上的多层次编码并受到听者目标选择的影响（Zion-Golumbic & Schroeder, 2012; Mesgarani & Chang, 2012）。

与言语中的层级结构相似，不同频率的神经振荡也趋向于以一种层级化的模式相互耦合。大脑内的低频振荡（如 theta 频带）可能反应了音节层面加工，而高频振荡（如 gamma 频带）更多表征了音素或发音特征等信息，频带间的相互耦合反应了远距离脑区信息交流以及协调全局神经网络的信息整合（Baltus & Herrman, 2016）。在 A1 中，Gamma 频带振幅随 theta 振荡的相位系统变化，theta 振幅还与 delta（1-2 Hz）相位耦合（Lakatos et al., 2007; Lakatos et al., 2005）。有趣的是这类效应受言语可懂度的影响，相比于倒放言语（无法理解的），加工自然言语（可理解的）时，听者的左侧额下区的 delta 频带和中央前回的 theta 频带才能够调节左侧听觉区域 25 Hz 振荡的相位活动（Park et al., 2015）。所以不同节律的神经振荡能够表征言语刺激中不同时间尺度的层级信息，通过相互协调对这些信息进行整合处理，完成听觉言语理解（Kayser et al., 2015）。

3.2 神经夹带连接内外节律

外部节律性刺激输入时，听者大脑会记录到与外部节律在相位上相对齐或相同频带下能量增大的现象（Kösem et al., 2018; Obleser & Kayser, 2019）。这些过程被认为可能是由于持续的神经振荡的相位被外部节律刺激所“重置”（reset）而产生的（Lakatos et al., 2009），我们通常将这种内部节律与外部节律的时间对齐现象称为神经夹带。研究者一般认为神经夹带的发生是基于神经系统自身具有的节律性活动，它们能够在缺乏外界连续性刺激输入的情况下维持活动，所以夹带能够在外部刺激消失后维持一段时间（Kösem et al., 2018; Tass et al., 1998）。常见的神经夹带计算方法包括外部刺激和大脑活动之间的相位相干性，以及以正向(如时间响应函数)或反向（如刺激重构)的方式连接大脑和刺激的回归模型（Fiedler et al., 2019; Fuglsang et al., 2017; Zhang & Ding, 2017）。所以神经夹带有时也被称为同步化（synchronization），或外部刺激为听觉言语时还被称为言语追随（speech tracking）反应。

在言语理解中，外部节律可能源于音节、词汇边界或其他声学线索。神经夹带能够通过这些外部节律特征完成语音分析，从连续的声音信号中提取离散的语言成分（Haegens & Golumbic, 2018; Obleser & Kayser, 2019）。经典神经夹带观点认为产生夹带的神经活动相位与言语中韵律或音节边界一致（Giraud & Poeppel, 2012; Peelle & Davis, 2012），例如通过 delta 节律振荡追踪韵律线索（Bourguignon et al., 2013）， theta 节律振荡反映音节和词汇结构（Doelling et al., 2014; Ding et al., 2016）。Luo 和 Poeppel(2007)通过记录听者在加工自然言语情况下的皮层脑磁图信号发现，大脑的 theta 节律振荡的相位模式稳定地追随口语句子中的音节节律。此研究还发现当言语与噪声相互嵌合后，外部言语的节律性丧失以及句子可懂度下降会破坏听者的神经夹带反应。除了其他声学刺激干扰外，言语本身的语速也会影响神经夹带，一旦语速过快，听者将难以跟上句子内容使夹带被中断（Ahissar et al., 2001）。有趣的是，对于言语内的物理声学特征的夹带是自动的，如睡眠期间也能记录到夹带（Ding & He, 2016; Makov et al., 2017）。但涉及句子内的语言学单元则需要言语被注意或者理解时才能够被夹带（Brodbeck et al., 2018）。虽然可理解性与语速快慢的研究反映神经夹带在言语处理中的作用，但此类研究仍旧存在一些问题。降低语音可理解性通常涉及刺激声学的变化，因此观察到的语音跟踪反应的差异可能与改变的声音输入有关（Ding & Simon, 2012; Kosem & van Wassenhove, 2017; Steinmetzger & Rosen, 2017）。所以未来研究中，探索言语理解与神经夹带的关系，需要仔细控制语音刺激的声学特性 。

在面对面的交谈中，听者言语理解受身体语言的影响（Morillon & Baillet, 2017; Morillon et al., 2014; Poeppel & Assaneo, 2020)。Park 等人（2016）发现这一过程同样涉及说话者运动

与听者神经活动的夹带过程。他们通过计算说话者口唇开合面积随时间变化模式与听者初级视听皮层和左侧运动区的神经活动，发现二者在 1Hz 上有显著的夹带效应，并且目标言语的理解程度能被夹带的同步性所预测。听者大脑对身体言语的夹带使听者能够实时利用运动区“模拟”的发声过程帮助听觉区预测即将输入的感觉刺激，进而促进言语理解（Morillon & Baillet, 2017; Morillon et al., 2014）。

神经夹带反映了节律信息对言语理解的影响。言语在时间结构上的规律性作为外部节律被听者感知时，大脑内部神经活动的重置使内部节律的变化模式与外部节律相似，这使得相近相位模式下的神经活动成为言语理解的理想环境（Haegens & Golumbic, 2018; Schroeder & Lakatos, 2009）。然而，神经夹带并非单一对外部节律信息的被动追随，它还受听者主观调控的影响。我们将从言语理解过程中涉及的几个自上而下的认知过程对神经夹带的影响，进一步探讨神经夹带在言语理解的作用。

4 自上而下调节神经夹带对言语理解的作用

神经夹带能根据听者当前的认知状态动态选择或增强与外部输入的同步性，便于大脑更有针对性的预测目标信息（Lakatos et al., 2019）。在言语理解过程中，自上而下的调控可能源于听者的选择性注意（Helfrich et al., 2019; Lakatos et al., 2013; Obleser & Kayser, 2019）、语法的先验知识（Ding et al., 2016; Ding et al., 2017）和言语语境产生的预期（Broderick et al., 2019）等认知过程。

嘈杂的声学环境使目标言语理解变得困难，选择性注意有助于放大注意刺激流与非注意刺激间的夹带差异，前者与神经活动的相位同步有利于获取更多的加工资源，后者则传递到了非最优相位阶段使其更容易被抑制，这有助于嘈杂环境中的言语理解（Calderone et al., 2014; Knudsen, 2018; Lavie, 1995; Zion-Golumbic & Schroeder, 2012）。多说话者的场景中，听者选择性注意单一说话人的言语内容时，听觉皮层区域（如：颞上回）和高层级的脑区（如：额下皮层，颞前部）都发现增强了神经振荡的振幅调制，高级皮层区域还表现出更明显的选择性增强对注意言语夹带的现象（Golumbic et al., 2013）。此外，选择性注意还有助于身体语言促进听觉言语理解过程。当听者更加注意说话者的口唇运动时，左侧运动皮层与口唇运动间的夹带增强，并且这种增强能直接预测言语理解的准确性（Park et al., 2016）。所以不同脑区间的神经夹带能够通过选择性注意建立起时间上的耦合，提高脑区间的信息整合的精确度。

言语理解过程需要通过语音特征检索对应的词汇信息，再基于听者先验的语法知识组合成短语和句子（Poeppel et al., 2008；Townsend and Bever, 2001; Phillips, 2003）。在排除

口语韵律和统计层面线索的影响下，研究者发现不同频率的皮层活动能够同时追随言语中单词、短语和句子等不同层次的抽象语言结构的时间进程（Ding et al., 2016）。不同时间尺度的言语单元的同步神经夹带可能预示着一种层级嵌入模式，即更小的言语单元表征嵌入在更高层次的言语单元表征下（Lerner et al., 2011; Christiansen et al., 2016; Poeppel et al., 2008），从而实现言语中不同层级信息间的及时整合(Ding et al., 2017; Ding et al., 2016)。当听者理解言语内容后，依据上下文产生的语境预期同样能够影响之后出现词汇的语音包络的神经夹带程度，即词汇的语义与上下文越接近时目标词汇的皮层脑电信号的神经夹带越强（Broderick et al., 2019）。这表明神经夹带还受到听者基于上下文预测的影响，能够最大限度地提高未来事件的可预测性，并精确地安排资源的分配时间（Henry et al., 2014)，从而促进即将到来的单词的初级编码阶段处理。这一机制还解释了为什么外部可预测的节律性刺激相比不可预测的非节律性刺激更容易被感知（Mathewson et al., 2010; Rohenkohl et al., 2012）。

神经夹带受听者自上而下认知过程的影响能够更好的表征复杂听觉环境中的节律信息，促进目标言语的理解。它可以作为一个“滤波器”，根据听者的选择性注意减弱或消除高级脑区在嘈杂环境中对非注意语音流的神经响应；还可以作为一个“增益器”，依据听者的预期来增强言语中相应成分的表征和加工；最后神经夹带可以作为一个“连接器”，根据听者已有的先验知识完成言语内不同层级间成分或跨脑区间信息的整合。所以听者的主动调控使言语理解过程中的关键信息具有更大的可能性处于神经元集群活动的最佳兴奋性水平，从而获得更多的加工资源。据此我们认为神经夹带可能为外部节律和内部节律提供了一座联系的“桥梁”。

5 讨论

谈及言语节律时，研究者往往从语音信号的时域层面进行讨论，如语音包络或声学单位的持续时间等。这些对象的起伏变化构成了听者对外部节律的感知，是构建可理解言语过程所必需的。外部节律能够帮助听者关注单词或它们的组成元素(如音素或音节)进而促进理解。听者大脑的神经振荡作为内部节律，能够表征与整合言语内不同层级信息。而神经夹带可能是实现言语理解中内外节律相互联系的关键。

5.1 外部节律促进言语理解

口语中的词汇间停顿、停顿位置等韵律结构节律影响听者在口语加工中可懂度的高低以及对歧义语境的结构分析，合适的韵律结构节律能够促进正确的言语理解，恢复难以理解的语义内容（Ghitza & Greenberg, 2009; Li & Yang, 2009, 2010)。而于不同的语速的语境

则改变听者对随后出现的音节判别乃至词汇数量感知（Dilley & Pitt, 2010; Bosker, Sjerps, et al., 2020; Reinisch, 2016）。此外，说话者在言语产生过程中伴随的同步运动行为能够通过视觉通道同时与言语信息传入听者大脑。这些非听觉的运动节律与言语节律之间的协同性能够帮助听者更好的捕捉目标言语内容（Bosker & Peeters, 2021; Poeppel & Assaneo, 2020）。所以言语理解得益于这些外部节律特征，它们不但能够帮助听者理解和降低处理成本，还能调节音素、词汇和句子层面的语音处理。

当以语音材料的持续时间为对象研究言语节律时，口语中声学单元的持续时间会改变说话者语速的感知。语速快慢通常是改变口语中元音音程百分比（the percentage of vocalic intervals, %V）和辅音音程的标准差（the standard deviation of consonantal intervals, deltaC）进而影响节律感知，但这一现象并非所有语种中都存在，如法语的语速快慢并不影响 deltaC 的变异系数（Dellwo, 2006；Dellwo et al., 2003）。所以不同语种中涉及语速变化是否能直接影响言语节律感知仍旧存在争议，这提示进行不同语言中涉及声学单元持续时间的研究必须关注语速的操控。

5.2 神经夹带——联系内外节律的可能机制

大脑神经活动的揭示使研究者认为，内部节律性神经振荡表征了言语信号，从而实现听者对信号中关键信息的加工，在音节感知、语义加工和句法理解方面得到了证实（Cason & Schoen, 2012; Kotz & Schmidt-Kassow, 2015; Schmidt-Kassow et al., 2013）。近些年，大量的研究发现大脑内的神经振荡在时间上可能存在与外部节律性刺激间的夹带现象（Kösem et al., 2018; Obleser & Kayser, 2019）。由于持续性神经活动的相位反映了神经元兴奋性的节律性波动，当夹带发生时，产生夹带效应的神经活动与外部刺激在时间上相互对齐时，能够实现稳定地调整对输入刺激的加工增益（Buzsaki & Draguhn, 2004; Lakatos et al., 2005; Vanthornhout et al., 2018）。所以我们认为神经夹带是实现言语理解中内外部节律相互联系的可能机制。

神经夹带现象广泛的存在于外部节律影响言语理解的过程中。它为大脑如何表征言语中不同层级信息提供了途径。对言语刺激的夹带并非发生于某一特定频段中，从反应声学特征的 Gamma 频段到语音的时域包络的 theta 频段，或是汉语中字，词和句中更低的 delta 频带，不同层级下的节律大脑都有相对应的神经振荡产生夹带（Giraud & Poeppel, 2012; Peelle & Davis, 2012）。神经夹带还说明了韵律结构节律或语境节律的建立对当下言语理解的影响可能是由于自身所具有的自我维持特性。即在事先输入的节律刺激变化后已经产生的夹带仍旧可以持续一段时间进而影响当前输入言语刺激的加工（Kösem et al.,

2018）。对身体言语的夹带有助于跨脑区间信息交流的锁时性，确保运动信息能够与言语信息精确的整合（Park et al., 2016）。

听者自上而下认知过程调节神经夹带为选择性注意、先验知识和预期在言语理解中的作用提供了生理层面解释。听者的选择性注意通过神经夹带使得高兴奋性的神经集群能够更为集中地表征目标刺激，进而提高目标言语的识别率（Calderone et al., 2014; Golumbic et al., 2013; Knudsen, 2018; Lavie, 1995）。反之，神经活动会阻碍无法对齐的感觉刺激的表征建立，因为它们会随机地放大或衰减信息（Lakatos et al., 2019）。先验的语法知识则通过同时对言语中不同层级单位的夹带实现层级间的整合的精确性（Ding et al., 2016; Ding et al., 2017）。而当听者理解上下文内容时，对于之后出现词汇的预期能够加强在加工词汇时的夹带强度，促进词汇的早期发音编码加工（Broderick et al., 2019）。所以我们认为神经夹带不单是一种被动相应外部节律性刺激的大脑活动，还能够根据听者的认知状态创建一个合适的当前言语理解的加工环境。它作为量化两种节律性活动一致性关系的指标，已成为描述外部言语与大脑之间双向关系的方法，允许研究者探究节律或听者认知过程是如何影响言语理解。

5.3 存在的挑战

长期以来关于大脑对感觉刺激的反应是否与内在的、持续的神经振荡有关一直存在争议（Doelling & Assaneo, 2021），大脑的神经夹带是否由神经振荡所产生也缺少直接证据。研究者需要严谨地判断结果中的夹带现象是由外部刺激和内在神经振荡之间的耦合产生，还是一连串刺激引起的一系列神经元诱发电位。在许多情况下，所谓的夹带可能只是来自于声音的规律性输入而引起的一系列神经元诱发响应，而非真正的神经振荡（Obleser & Kayser, 2019; Poeppel & Assaneo, 2020）。

随着非侵入性脑刺激技术的发展，研究者不在局限于被动记录听者的大脑活动，而开始采用外加干预的形式探究神经振荡对言语理解的影响。经颅交流电刺激（tACS）相比于经颅磁刺激（TMS）是一种完全无声的刺激手段，可以排除任务中实验外声音的干扰（Feher et al., 2017; Raco et al., 2016）。当听者在加工言语刺激时对其颞叶区施加 tACS，干扰 theta 频段神经振荡的活动，结果发现受到干扰的神经活动导致神经夹带破坏，并表现出言语可懂度成绩的下降（Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018）。相反，如果将言语刺激的包络作为电刺激模态时，对头皮的刺激能够提高听者在噪声环境下的言语理解能力（Keshavarzi & Reichenbach, 2020）。tACS 在刺激频率和相位上与节律性听觉刺激对齐，有助于听皮层对连续听觉事件的感知（Wilsch et al., 2018; Zoefel et al.,

2018）。未来研究中，无论是控制神经振荡与外部刺激的同步化还是去同步化，都有助于提供神经振荡在言语知觉中的作用的更直接证据。并且这种对大脑振荡的“实验性”影响允许通过检查其功能结果来确定大脑振荡是否因果驱动大脑功能，而不是一种副现象活动（Vosskuhl et al., 2018）。

6 总结

听觉言语理解涉及多尺度的内外部节律共同参与。我们首先通过韵律结构节律、语境节律、和说话者身体语言节律三个常见的外部节律，揭示了外部节律能够影响听觉言语理解。其次，我们描述了听者内部神经振荡与神经夹带现象在言语理解过程中的作用。最后，我们根据神经夹带受听者自上而下认知过程的影响，探讨神经夹带可能是联系内外节律的关键机制。

参考文献

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of Multimovement coordination: Sensorimotor mechansims in speech motor programming. *Journal of Motor Behavior*, *16*(2), 195-231. https://doi.org/https://doi.org/10.1080/00222895.1984.10735318

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367-13372. https://doi.org/10.1073/pnas.201400998

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390-398. https://doi.org/10.1016/j.tics.2012.05.003

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-Term Temporal Tracking of Speech Rate Affects Spoken-Word Recognition. *Psychological Science*, *25*(8), 1546-1553. https://doi.org/10.1177/0956797614533705

Baltus, A., & Herrman, C. S. (2016). The importance of individual frequencies of endogenous brain oscillations for auditory cognition - A short review. *Brain Research*, *1640*, 243-250. https://doi.org/10.1016/j.brainres.2015.09.030

Bishop, G. H. (1933). Cyclic changes in excitability of the optic pathway of the rabbit. *American Journal of Physiology*, *103*(1), 213-224. https://doi.org/https://doi.org/10.1152/ajplegacy.1932.103.1.213

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention Perception & Psychophysics*, *79*(1), 333-343. https://doi.org/10.3758/s13414-016-1206-4

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: behavioural evidence from rate normalisation. *Language Cognition and Neuroscience*, *33*(8), 955-967. https://doi.org/10.1080/23273798.2018.1439179

Bosker, H. R., & Peeters, D. (2021). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B-Biological Sciences*, *288*(1943). https://doi.org/10.1098/rspb.2020.2419

Bosker, H. R., Peeters, D., & Holler, J. (2020). How visual cues to speech rate influence speech perception. *Quarterly Journal of Experimental Psychology*, *73*(10), 1523-1536. https://doi.org/10.1177/1747021820914564

Bosker, H. R., Sjerps, M. J., & Reinisch, E. (2020). Temporal contrast effects in human speech perception are immune to selective attention. *Scientific Reports*, *10*(1), Article 5607. https://doi.org/10.1038/s41598-020-62613-8

Bourguignon, M., De Tiege, X., Op de Beeck, M., Ligot, N., Paquier, P., Van Bogaert, P., . . . Jousmaki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314-326. https://doi.org/10.1002/hbm.21442

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *Plos Biology*, *15*(2), Article e2001665. https://doi.org/10.1371/journal.pbio.2001665

Bridwell, D. A., Henderson, S., Sorge, M., Plis, S., & Calhoun, V. D. (2018). Relationships between alpha oscillations during speech preparation and the listener N400 ERP to the produced speech. *Scientific Reports*, *8*, Article 12838. https://doi.org/10.1038/s41598-018-31038-9

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic

Representations of Continuous Speech. *Current Biology*, *28*(24), 3976-+. https://doi.org/10.1016/j.cub.2018.10.042

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current Biology*, *28*(5), 803-+. https://doi.org/10.1016/j.cub.2018.01.080

Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic context enhances the early auditory encoding of natural speech. *The Journal of Neuroscience*, *39*(38), 7564-7575. https://doi.org/10.1523/jneurosci.0584-19.2019

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180. https://doi.org/10.1159/000261913

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189-1196. https://doi.org/10.3758/s13423-011-0167-9

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926-1929. https://doi.org/10.1126/science.1099745

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300-309. https://doi.org/10.1016/j.tics.2014.02.005

Cason, N., & Schoen, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652-2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Christiansen, M. H., & Chater, N. (2015). The now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/s0140525x1500031x

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *72*, 52-65. https://doi.org/10.1016/j.wocn.2018.11.002

Dauer, R. M. (1983). Stress-Timing and Syllable-Ttimg Reanalyzed. *Journal of Phonetics*, *11*(1), 51-62. https://doi.org/10.1016/s0095-4470(19)30776-4

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th linguistic colloquium* (pp. 231–241).

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. *International Congress of Phonetic Sciences*, Barcelona/Spain. https://doi.org/10.5167/uzh-111779

Di Liberto, G. M., Wong, D., Melnik, G. A., & de Cheveigne, A. (2019). Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *Neuroimage*, *196*, 237-247. https://doi.org/10.1016/j.neuroimage.2019.04.037

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274-294. https://doi.org/10.1016/j.jml.2010.06.003

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294-311. https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or

Disappear. *Psychological Science*, *21*(11), 1664-1670. https://doi.org/10.1177/0956797610384743

Ding, N., & He, H. (2016). Rhythm of Silence. *Trends in Cognitive Sciences*, *20*(2), 82-84. https://doi.org/10.1016/j.tics.2015.12.006

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158-+. https://doi.org/10.1038/nn.4186

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, *81*, 181-187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *The Journal of Neuroscience*, *38*(5), 1178-1188. https://doi.org/10.1523/jneurosci.2606-17.2017

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78-89. https://doi.org/10.1152/jn.00297.2011

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761-768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Doelling, K. B., & Assaneo, M. F. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLOS Biology*, *19*(5), e3001234. https://doi.org/10.1371/journal.pbio.3001234

Farbood, M. M., Marcus, G., & Poeppel, D. (2013). Temporal Dynamics and the Identification of Musical Key. *Journal of Experimental Psychology-Human Perception and Performance*, *39*(4), 911-918. https://doi.org/10.1037/a0031087

Feher, K. D., Nakataki, M., & Morishima, Y. (2017). Phase-Dependent Modulation of Signal Transmission in Cortical Networks through tACS-Induced Neural Oscillations. *Frontiers in Human Neuroscience*, *11*, Article 471. https://doi.org/10.3389/fnhum.2017.00471

Fiedler, L., Woestmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage*, *186*, 33-42. https://doi.org/10.1016/j.neuroimage.2018.10.057

Fuglsang, S. A., Dau, T., & Hjortkjaer, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage*, *156*, 435-444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers in Human Neuroscience*, *8*, Article 777. https://doi.org/10.3389/fnhum.2014.00777

Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: vision, rhythm, cooperation. *Trends in Cognitive Sciences*, *18*(10), 543-553. https://doi.org/10.1016/j.tics.2014.06.004

Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence.

Phonetica, *66*(1-2), 113-126. https://doi.org/10.1159/000208934

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511-517. https://doi.org/10.1038/nn.3063

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., . . . Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron*, *77*(5), 980-991. https://doi.org/10.1016/j.neuron.2012.12.037

Haegens, S., & Golumbic, E. Z. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*, 150-165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Helfrich, R. F., Breska, A., & Knight, R. T. (2019). Neural entrainment and network resonance in support of top-down guided attention. *Current Opinion in Psychology*, *29*, 82-89. https://doi.org/10.1016/j.copsyc.2018.12.016

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(41), 14935-14940. https://doi.org/10.1073/pnas.1408741111

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639-652. https://doi.org/10.1016/j.tics.2019.05.006

Iani, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110-121. https://doi.org/10.1016/j.jml.2017.05.004

Jadoul, Y., Ravignani, A., Thompson, B., Filippi, P., & de Boer, B. (2016). Seeking Temporal Predictability in Speech: Comparing Statistical Approaches on 18 World Languages. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00586

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200-206. https://doi.org/10.1016/j.tics.2012.03.002

Kayser, C. (2019). Evidence for the Rhythmic Perceptual Sampling of Auditory Scenes. *Frontiers in Human Neuroscience*, *13*, https://doi.org/10.3389/fnhum.2019.00249

Kayser, C., Wilson, C., Safaai, H., Sakata, S., & Panzeri, S. (2015). Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. *Journal of Neuroscience*, *35*(20), 7750-7762. https://doi.org/10.1523/jneurosci.0268-15.2015

Keshavarzi, M., & Reichenbach, T. (2020). Transcranial Alternating Current Stimulation With the Theta-Band Portion of the Temporally-Aligned Speech Envelope Improves Speech-in-Noise Comprehension. *Frontiers in Human Neuroscience*, *14*, https://doi.org/10.3389/fnhum.2020.00187

Knudsen, E. I. (2018). Neural Circuits That Mediate Selective Attention: A Comparative Perspective. *Trends in Neurosciences*, *41*(11), 789-805. https://doi.org/10.1016/j.tins.2018.06.006

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*(18), 2867-2875. https://doi.org/10.1016/j.cub.2018.07.023

Kösem, A., & van Wassenhove, V. (2016). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language Cognition and Neuroscience*, *32*(5), 536-544. https://doi.org/10.1080/23273798.2016.1238495

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896-910. https://doi.org/10.1016/j.tics.2018.08.002

Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, *68*, 48-60. https://doi.org/10.1016/j.cortex.2015.02.021

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392-399. https://doi.org/10.1016/j.tics.2010.06.005

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Ladefoged, P. (1975). *A Course in Phonetics*. Harcourt Brace Jovanovich College.

Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. *Current Biology*, *29*(18), R890-R905. https://doi.org/10.1016/j.cub.2019.07.075

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. *Neuron*, *64*(3), 419-430. https://doi.org/10.1016/j.neuron.2009.10.014

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, *94*(3), 1904-1911. https://doi.org/10.1152/jn.00263.2005

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology-Human Perception and Performance*, *21*(3), 451-468. https://doi.org/10.1037/0096-1523.21.3.451

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive Windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906-2915. https://doi.org/10.1523/jneurosci.3684-10.2011

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese Sentences. *Neuroscience*, *158*(4), 1416-1425. https://doi.org/10.1016/j.neuroscience.2008.10.065

Li, W., & Yang, Y. (2010). Perception of Chinese poem and its electrophysiological effects. *Neuroscience*, *168*(3), 757-768. https://doi.org/10.1016/j.neuroscience.2010.03.069

Li, W., Zhang, H., Zheng, Z., & Li, X. (2019). Prosodic phrase priming during listening to Chinese ambiguous phrases in different experimental tasks. *Journal of Neurolinguistics*, *51*, 135-150. https://doi.org/10.1016/j.jneuroling.2019.02.003

Li, X., & Ren, G. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia*, *50*(8), 1882-1894. https://doi.org/10.1016/j.neuropsychologia.2012.04.013

Li, X., Shao, X., Xia, J., & Xu, X. (2019). The cognitive and neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension. *Journal of Neurolinguistics*, *49*, 155-167. https://doi.org/10.1016/j.jneuroling.2018.05.004

Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*, 377-401. https://doi.org/10.1177/00238309000430040301

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in

human auditory cortex. *Neuron*, *54*(6), 1001-1010. https://doi.org/10.1016/j.neuron.2007.06.004

Luo, Y., Duan, Y., & Zhou, X. (2015). Processing rhythmic pattern during Chinese sentence reading: An eye movement study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01881

Luo, Y., & Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage*, *49*(3), 2836-2849. https://doi.org/10.1016/j.neuroimage.2009.10.008

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep Disrupts High-Level Speech Parsing Despite Significant Basic Auditory Processing. *Journal of Neuroscience*, *37*(32), 7772-7781. https://doi.org/10.1523/jneurosci.0168-17.2017

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019). How the Tracking of Habitual Rate Influences Speech Perception. *Journal of Experimental Psychology-Learning Memory and Cognition*, *45*(1), 128-138. https://doi.org/10.1037/xlm0000579

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition*, *115*(1), 186-191. https://doi.org/10.1016/j.cognition.2009.11.010

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233-U118. https://doi.org/10.1038/nature11020

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(42), E8913-E8921. https://doi.org/10.1073/pnas.1705373114

Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, *5*, Article 5255. https://doi.org/10.1038/ncomms6255

Morris, D. J., & Klerke, S. (2016). Machine classification of P1-N1-P2 responses elicited with a gated syllable. *The Journal of the Acoustical Society of America*, *140*(4), 3155-3155. https://doi.org/https://doi.org/10.1121/1.4969899

Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., & Tessendorf, S. (2013). Body-language-communication: . In *An international handbook on multimodality in human interaction* (Vol. 1, pp. 131-232). De Gruyter Mouton.

Nooteboom, S. (1997). The prosody of speech: melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of the phonetic sciences* (Vol. 5, pp. 640-673). Blackwell Publishers.

O'Brien, G. E., Gijbels, L., & Yeatman, J. D. (2020). Context effects on phoneme categorization in children with dyslexia. *Journal of the Acoustical Society of America*, *148*(4), 2209-2222. https://doi.org/10.1121/10.0002181

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife*, *5*. https://doi.org/10.7554/elife.14521

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, *25*(12), 1649-1653. https://doi.org/10.1016/j.cub.2015.04.049

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, https://doi.org/10.3389/fpsyg.2012.00320

Pike, K. L. (1945). The Intonation of American English.

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention Perception & Psychophysics*, *78*(1), 334-345. https://doi.org/10.3758/s13414-015-0981-7

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322-334. https://doi.org/10.1038/s41583-020-0304-4

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071-1086. https://doi.org/10.1098/rstb.2007.2160

Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., & Narayanan, S. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, Article 100921. https://doi.org/10.1016/j.wocn.2019.100921

Raco, V., Bauer, R., Tharsan, S., & Gharabaghi, A. (2016). Combining TMS and tACS for Closed-Loop Phase-Dependent Modulation of Corticospinal Excitability: A Feasibility Study. *Frontiers in Cellular Neuroscience*, *10*, https://doi.org/10.3389/fncel.2016.00143

Ramus, F. (2002). *Acoustic correlates of linguistic rhythm: Perspectives* Proc Speech Prosody, Aix-en-Provence.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292. https://doi.org/10.1016/s0010-0277(99)00058-x

Reichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, *14*(4), 180-190. https://doi.org/10.1016/j.tics.2010.01.008

Reinisch, E. (2016). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention Perception & Psychophysics*, *78*(4), 1203-1217. https://doi.org/10.3758/s13414-016-1067-x

Riecke, L., Formisano, E., Sorger, B., Baskent, D., & Gaudrain, E. (2018). Neural Entrainment to Speech Modulates Speech Intelligibility. *Current Biology*, *28*(2), 161-169. https://doi.org/10.1016/j.cub.2017.11.033

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870-882. https://doi.org/10.1016/j.tics.2018.08.003

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies*, *73*, 79.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal Expectation Improves the Quality of Sensory Information. *Journal of Neuroscience*, *32*(24), 8424-8428. https://doi.org/10.1523/jneurosci.0804-12.2012

Schmidt-Kassow, M., Roncaglia-Denissen, M. P., & Kotz, S. A. (2013). Speech Rhythm Facilitates Syntactic Ambiguity Resolution: ERP Evidence. *Figshare*. https://doi.org/https://doi.org/10.1371/journal.pone.0056000

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9-18. https://doi.org/10.1016/j.tins.2008.09.012

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and

visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106-113. https://doi.org/10.1016/j.tics.2008.01.002

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., . . . Gao, J.-H. (2019). The Cortical Maps of Hierarchical Linguistic Structures during Speech Perception. *Cerebral Cortex*, *29*(8), 3232-3240. https://doi.org/10.1093/cercor/bhy191

Steinmetzger, K., & Rosen, S. (2017). Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech. *Neuropsychologia*, *95*, 173-181. https://doi.org/10.1016/j.neuropsychologia.2016.12.003

Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews-Cognitive Science*, *11*(1), Article e1517. https://doi.org/10.1002/wcs.1517

Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., . . . Freund, H. J. (1998). Detection of n : m phase locking from noisy data: Application to magnetoencephalography. *Physical Review Letters*, *81*(15), 3291-3294. https://doi.org/10.1103/PhysRevLett.81.3291

Thut, G., Schyns, P. G., & Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Psychology*, *2*, Article 170. https://doi.org/10.3389/fpsyg.2011.00170

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapic, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93-118. https://doi.org/10.1515/lp-2013-0005

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *Jaro-Journal of the Association for Research in Otolaryngology*, *19*(2), 181-191. https://doi.org/10.1007/s10162-018-0654-z

Vosskuhl, J., Strueber, D., & Herrmann, C. S. (2018). Non-invasive Brain Stimulation: A Paradigm Shift in Understanding Brain Oscillations. *Frontiers in Human Neuroscience*, *12*, Article 211. https://doi.org/10.3389/fnhum.2018.00211

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939-950. https://doi.org/10.3758/bf03193621

Wang, M., Kong, L., Zhang, C., Wu, X., & Li, L. (2018). Speaking rhythmically improves speech recognition under "cocktail-party" conditions. *Journal of the Acoustical Society of America*, *143*(4), EL255-EL259. https://doi.org/10.1121/1.5030518

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, *63-64*, 38-54. https://doi.org/10.1016/j.specom.2014.04.003

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665-679. https://doi.org/10.1016/j.jml.2011.12.010

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *Neuroimage*, *172*, 766-774. https://doi.org/10.1016/j.neuroimage.2018.01.038

Wu, C., Cao, S., Wu, X., & Li, L. (2013). Temporally pre-presented lipreading cues release speech from informational masking. *Journal of the Acoustical Society of America*, *133*(4), EL281-EL285. https://doi.org/10.1121/1.4794933

Zhang, W., & Ding, N. (2017). Time-domain analysis of neural tracking of hierarchical linguistic structures. *Neuroimage*, *146*, 333-340. https://doi.org/10.1016/j.neuroimage.2016.11.016

Zion-Golumbic, E., & Schroeder, C. E. (2012). Attention modulates 'speech-tracking' at a cocktail party. *Trends in Cognitive Sciences*, *16*(7), 363-364. https://doi.org/10.1016/j.tics.2012.05.004

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Current Biology*, *28*(3), 401-408. https://doi.org/10.1016/j.cub.2017.11.071

方岚, 郑苑仪, 金晗, 李晓庆, 杨玉芳, & 王瑞明. (2021). 口语句子的韵律边界：窥探言语理解的秘窗. *心理科学进展*, *29*(3), 425-437. https://doi.org/https://dx.doi.org/10.3724/SP.J.1042.2021.00425

杨玉芳. (2021). *语言理解——认知过程和神经基础*. 科学出版社.

殷融. (2020). “动手不动口”: 手部动作与语言进化的关系. *心理科学进展*, *28*(7), 1141-1155. https://doi.org/10.3724/SP.J.1042.2020.01141

于泽, 韩玉昌, & 任桂琴. (2010). 韵律在语言加工中的作用及其神经机制. *心理科学进展*, *18*(3), 420-425.

The role of rhythm in auditory speech understanding

Abstract: Speech understanding is a mental process in which the listener receives external speech input and acquires meaning. In daily communication, speech comprehension is influenced by multi-scale rhythmic information, which usually includes the rhythm of prosodic structure, the rate of context, and the speaker's body language. They alter the listeners' phoneme discrimination, word perception, and speech intelligence in auditory speech understanding. Internal rhythms are neural oscillations in the brain, which can represent the hierarchical characteristics of external speech input at different time scales. The neural entrainment of external rhythmic stimulus and internal neural activity can optimize the brain's processing of speech stimulus and further enhance the internal representation of target speech by the top-down modulation of the listener's cognitive process. We think that it may be the key mechanism to build the interrelationship between internal and external rhythms and jointly affect speech understanding. The discovery of its mechanism can provide a window for the study of speech, which is a complex sequence with structural rules on multi-level time scales.

Keywords: rhythm, speech understanding, neural oscillation, neural entrainment, top-down modulation